MOLECULAR AND CELLULAR BIOLOGY, May 1989, p. 2279–2283
0270-7306/89/052279-05$02.00/0

Vol. 9, No. 5

Primary Structure of a Human Mitochondrial Protein Homologous to the Bacterial and Plant Chaperonins and to the 65-Kilodalton Mycobacterial Antigen

SATISH JINDAL, ANIL K. DUDANI, BHAG SINGH, CALVIN B. HARLEY, AND RADHEY S. GUPTA*

Department of Biochemistry, McMaster University, Hamilton, Ontario L8N 3Z5, Canada

Received 19 September 1988/Accepted 9 February 1989

The complete cDNA for a human mitochondrial protein designated P1, which was previously identified as a microtubule-related protein, has been cloned and sequenced. The deduced amino acid sequence of P1 shows strong homology (40 to 50% identical residues and an additional 20% conservative replacements) to the 65-kilodalton major antigen of mycobacteria, to the GroEL protein of *Escherichia coli*, and to the ribulose 1,5-bisphosphate carboxylase-oxygenase (rubisco) subunit binding protein of plant chloroplasts. Similar to the case with the latter two proteins, which have been shown to act as chaperonins in the posttranslational assembly of oligomeric protein structures, it is suggested that P1 may play a similar role in mammalian cells. The observed high degree of homology between human P1 and mycobacterial antigen also suggests the possible involvement of this protein in certain autoimmune diseases.

Our earlier studies with mutants of Chinese hamster ovary (CHO) cells selected for resistance to the microtubule (MT) inhibitor podophyllotoxin showed that a large number of these mutants involved specific electrophoretic alteration in a major protein designated P1 (M_r, ≈63 kilodaltons [kDa]) (6, 7). The genetic lesion in these mutants appears to be related to the cellular action of the drug, since podophyllotoxin-resistant mutants exhibit highly specific cross-resistance and collateral sensitivity to other MT inhibitors, such as colchicine, nocodazole, and taxol, and show reduced binding of the drug in cell extracts (6, 7). Immunofluorescence studies show that in interphase cells of vertebrate and invertebrate species, P1 antibody stains mitochondria, which show specific association with MTs (5, 7). Subfractionation of rat mitochondria has localized P1 to the matrix compartment (4). To help understand the cellular function of P1, cloning and sequencing of P1 cDNA from human cells was undertaken. The P1 sequence reported here shows extensive sequence and structural homology to a family of bacterial and plant proteins, termed chaperonins (8), which are involved in facilitating the posttranslational assembly of oligomeric protein complexes (1, 3, 8), as well as to the 65-kDa major antigenic protein of mycobacterial species (15, 18–20). The observed high degree of sequence and structural similarity between these proteins strongly indicates that P1 is the human homolog of this evolutionarily highly conserved group of proteins.

Isolation of P1-specific clones from λgt11 libraries. We have previously demonstrated that our antibodies to P1 cross-react only with the P1 protein in one- and two-dimensional immunoblots of proteins from CHO and human cells (5, 7). In view of this, a λgt11 cDNA library constructed from poly(A)$^+$ RNA from human HL-60 cells was screened by using P1 antiserum that had been preadsorbed to *Escherichia coli* proteins to remove any cross-reacting antibodies (11). From a total of 5×10^5 recombinant bacteriophage, eight P1-reactive clones, with inserts ranging from about 1.0 to 1.4 kilobase pairs (kbp), were identified. One clone, λ22a, containing a 1.4-kbp insert, was chosen for further studies. λ22a phage DNA was lysogenized in *E. coli* BNN103, and cell extracts from the lysogen were examined for the presence of a β-galactosidase–P1 fusion protein (Fig. 1a). In control cell lysates, a major band with the same mobility as β-galactosidase (M_r, ≈114 kDa) was observed. In contrast, BNN103 lysogenized with λ22a contained an apparent fusion protein band of higher M_r (≈150 kDa) which cross-reacted with the P1 antibody. In the same experiment, an extract from CHO cells showed specific cross-reactivity with a single protein of M_r 63,000, as expected.

To confirm that λ22a cDNA contained P1-specific sequences, the 150-kDa fusion protein was used to affinity purify antibodies from the original polyclonal serum. The purified antibody showed specific cross-reaction with a single protein of M_r 63,000 in CHO extracts, which was identical to that observed by using the original P1 antibody (Fig. 1b). Furthermore, the affinity-purified antibody also showed specific binding to mitochondria in immunofluorescence experiments, as was observed with the polyclonal antibody (results not shown).

In Northern (RNA) blot analysis of total RNA from either CHO or human cells, the 1.4-kbp cDNA insert from λ22a hybridized to a single mRNA of approximately 2.3 kbp (Fig. 1c). Nucleotide sequence analysis of the 1.4-kbp insert showed that it consisted of 1,485 bp, including a 3′ poly(A) tail, which indicated that it represented the 3′ end of P1 cDNA. Full-length cDNA was obtained from another HL-60 cDNA library, screened by using the 1.4-kbp probe. The clones obtained contained two *Eco*RI fragments, of approximately 1.4 and 0.8 kbp, of which only the 1.4-kbp fragment hybridized with the λ22a probe. Both fragments hybridized to an mRNA of 2.2 to 2.3 kbp in Northern blots (results not shown). One of these clones, λC5, was chosen for further investigations.

Nucleotide sequence of P1 cDNA. The complete nucleotide sequences of λ22a and λC5 were determined in both directions by using nested sets of deletions and oligonucleotide primers (Fig. 2). The cDNA insert in λC5 consisted of 2,242 nucleotides (nt) (−45 to 2197, where the A of the initiation codon is numbered 1), with an internal *Eco*RI site at nt 712. The nucleotide sequence of λ22a was identical to the 1,485 nt

* Corresponding author.

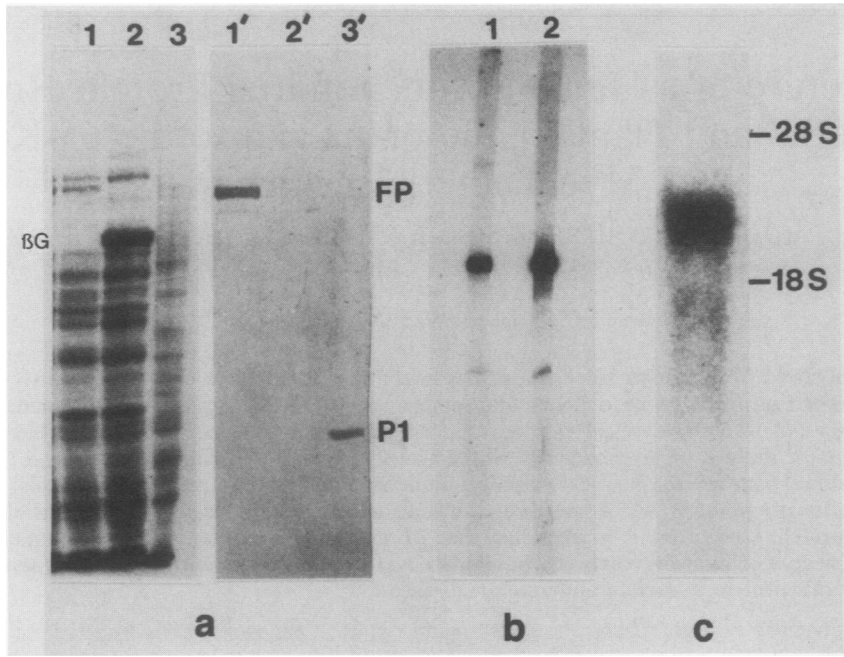

FIG. 1. Expression of recombinant β-galactosidase–P1 fusion protein in *E. coli*. (a) BNN103 cells lysogenized with either recombinant (λ22a) (lanes 1 and 1′) or nonrecombinant λgt11 (lanes 2 and 2′) were heat heat-induced, and cell extracts were analyzed on 7.5% sodium dodecyl sulfate-polyacrylamide gel electrophoresis, followed by staining with Coomassie blue (lanes 1 to 3) or immunoblotting with a 1:500 dilution of P1 antibody (lanes 1′ to 3′). Lanes 3 and 3′, CHO cell extract used as a reference for P1 antibody cross-reactivity. Positions of fusion protein (FP), P1, and β-galactosidase (βG) are marked. (b) Confirmation of λ22a clone by antibody selection. Antibodies from the P1 antiserum which bound to the 150-kDa fusion protein were isolated, and their reactivities with proteins from CHO cell extracts were compared. Lanes: 1, affinity-purified antibody; 2, original nonpurified antibody. (c) Northern blot analysis of P1 transcript in HeLa cells, using the λ22a (1.4-kbp) cDNA probe. Positions of 28S and 18S rRNA are marked.

downstream of the internal *Eco*RI site in clone λC5. The DNA sequence contains a single long open reading frame of 1,764 nt starting at the first base (nt −45). The termination triplet after 1,764 nt is followed by several additional termination signals in the same frame. At the 3′ end of the sequence, a 15-nt poly(A) tail is observed, preceded by a polyadenylation consensus sequence AATAAA present 15 nt upstream. The first ATG in the sequence (nt 1 to 3) likely constitutes the P1 initiation codon. It is flanked by sequence commonly found in eucaryotic mRNA translation initiation sites (G at nt −3 and −6, C at nt −5, −7, and −8, and absence of T between nt −17 and −1; 12), and the 26 amino acids encoded downstream of this ATG are characteristic of leader sequences from mitochondrial proteins (presence of Ser, Thr, and basic but not acidic residues) (16). The presence of a mitochondrial targeting sequence in the deduced sequence is in accordance with the subcellular localization of P1, and its size (M_r, 3.1 kDa) is consistent with the difference (≈3 kDa) between precursor and mature forms of P1 (4). Thus, human P1 mRNA encodes a 573-residue (M_r, 61,049; predicted pI, 6.3) precursor of P1. If the mitochondrial leader sequence is cleaved after Tyr-26, then mature P1 consists of 547 amino acids (M_r, 57,939; predicted pI, 6.0).

Homology of P1 to the chaperonin family of proteins. The deduced amino acid sequence of P1 shows a high degree of similarity to that of the GroEL protein of *E. coli* (8), the 65-kDa major antigen of mycobacteria (*Mycobacterium leprae*, *M. tuberculosis*, and *M. bovis* BCG) (15, 18–20), and to the rubisco (ribulose 1,5-bisphosphate carboxylase-oxygenase) large-subunit binding protein of plant chloroplasts (8) (Fig. 3A). Similarity between these proteins begins at residue 27 of P1, which supports the notion that Met-1 to Tyr-26 evolved as a leader sequence for mitochondrial targeting. The degree of sequence conservation between any two of these proteins is very similar: ≈40 to 50% of residues are identical, and an additional ≈20% are conservative replacements. In the multiple-sequence alignment, 28% of the residues are identical among all four proteins. The alignment score of any pair is about 50 standard deviations greater than the mean alignment score of corresponding randomized sequences. This high degree of statistical significance indicates that the similarity between these proteins represents homology by common descent. Given the long period since common ancestry of chloroplasts, mitochondria, and bacteria (>1 billion years; 17) and the high degree of homology, these proteins likely carry out a similar, primitive, and important function in these divergent species. Alignment of the predicted secondary structures of these proteins shows similar patterns of α helix interrupted by β sheet (Fig. 3B), which indicates that the overall structure of these proteins in various species is highly conserved.

The bacterial and plant proteins homologous to P1 display a number of common characteristics (8), including (i) high abundance, (ii) induction by heat shock, (iii) homo-oligomeric structure of either 7 or 14 subunits which reversibly dissociate in the presence of Mg^{2+} and ATP, (iv) ATPase activity, and (v) an essential but transient role in the correct folding and assembly of oligomeric protein structures. The last of these characteristics has led Ellis and co-workers (1, 8) to propose that these proteins, which are not part of the assembled complex, function as molecular chaperones within cells, i.e., proteins whose main function is to assist

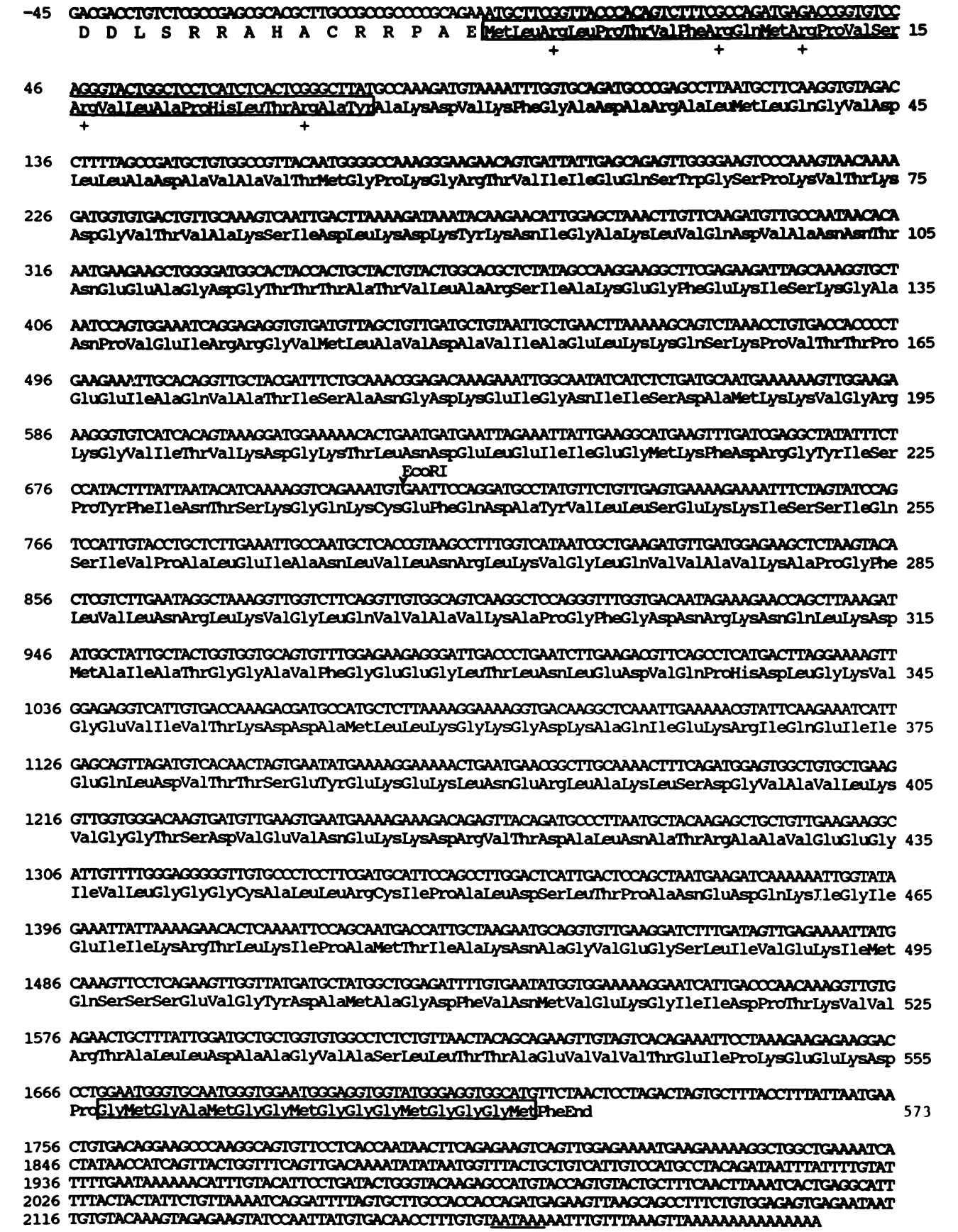

FIG. 2. Nucleotide and deduced amino acid sequences of the human P1. Numbers on the left refer to the nucleotide sequence relative to coordinate 1 at the beginning of the putative initiation codon. The amino acid sequence is numbered starting with 1 at the same point. The 5′ extension of this reading frame is shown in one-letter code. The position of the internal EcoRI site (nt 712), which marks the beginning of the λ22a sequence, is indicated. The polyadenylation signal 15 nt from the A tail at the 3′ end is underlined. The putative mitochondrial targeting sequence at the N-terminal end and a keratinlike (9) amino acid sequence at the C-terminal end containing repeats of Gly-Gly-Met are boxed. Positively charged amino acids in the leader sequence are identified (+).

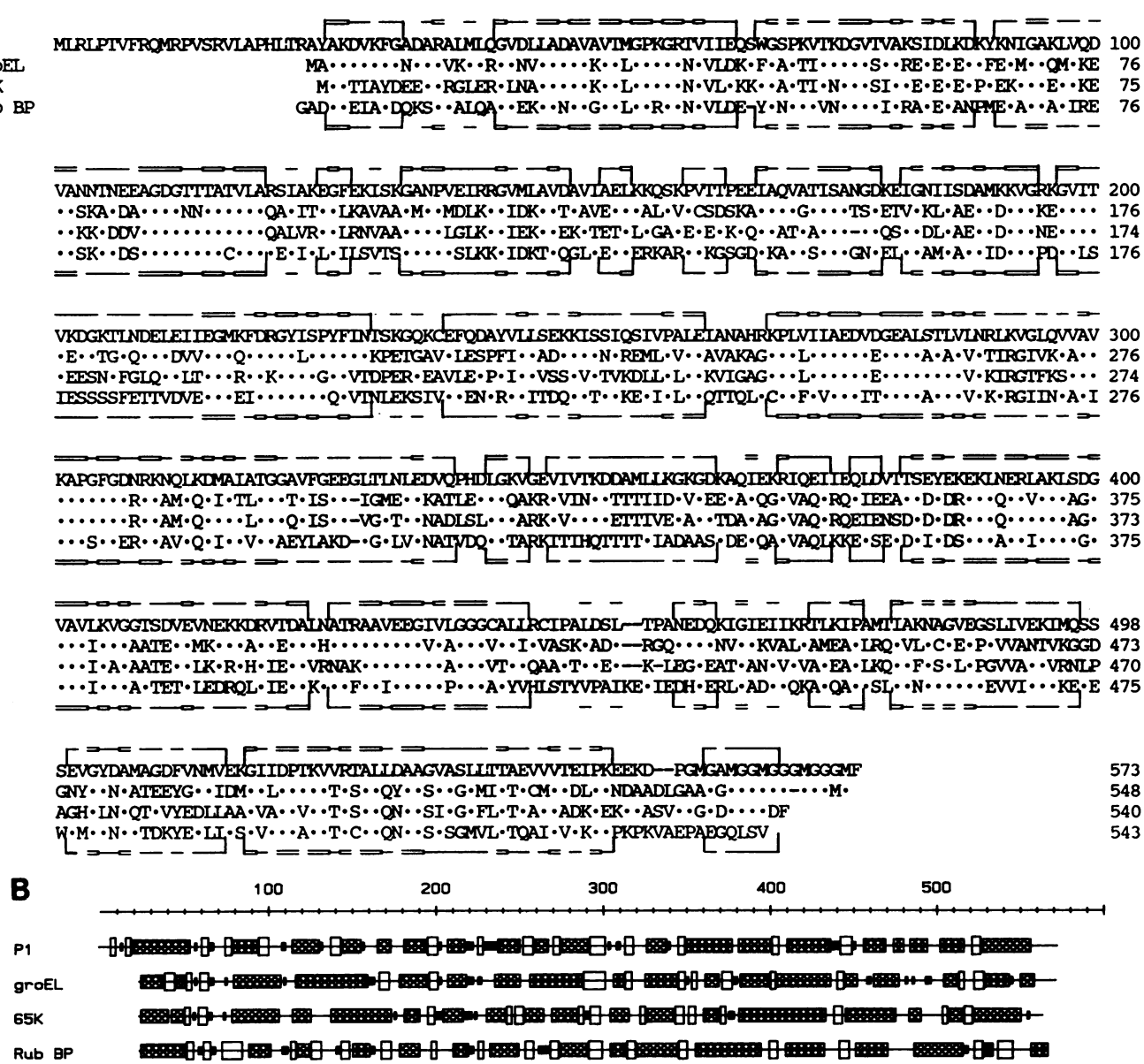

FIG. 3. (A) Multiple-sequence alignment of P1, GroEL (8), the 65-kDa major antigen from *M. bovis* BCG (20) (65K), and the rubisco large-subunit binding protein (Rub BP) from wheat (*Triticum aestivum*) (8). Residues identical to the corresponding P1 sequence are indicated by dots. Positions with identical residues in three or all four proteins are boxed with single or double lines, respectively. To indicate conservative changes based on amino acid similarity or common codon replacements, an ad hoc grouping of residues was used. Positions in which all four amino acids belong to one of the following groups are also boxed with a single line: (F, I, L, M, V), (F, Y, W), (H, N, Q), (D, E, G, N, Q, S, T), (K, R), or (A, P, G). Gaps in the sequence alignment are indicated by dashes. (B) Comparison of the secondary-structure predictions of P1, GroEL, *M. bovis* 65-kDa protein (65K), and the rubisco large-subunit binding protein (Rub BP). Structures predicted by the method of Garnier et al. (2) (α helix [▩], β sheet [□], and turns [■] are aligned as shown in panel A but without gaps.

other polypeptides to maintain or to assume conformations that permit their correct assembly into oligomeric structures. The bacterial-chloroplast-mitochondrial class of molecular chaperones was termed the chaperonin family (8). Similar to its bacterial and plant homologs, P1 is an abundant cellular protein which, in its native form, seems to exist as a homo-oligomer of seven subunits (approximate M_r of 440,000, as determined by gel filtration and sucrose density gradient centrifugation) and shows ATPase activity (unpublished results). These observations, together with the observed extensive sequence and structural similarity of P1 to the bacterial and plant chaperonins, strongly indicate that the mammalian P1 protein is a member of this highly conserved family of proteins. Although this is the first report regarding cloning and sequencing of the chaperonin gene from mammalian cells, McMullin and Hallberg (13, 14) have previously identified a 58-kDa heat shock protein (hsp58) in the mitochondria of *Tetrahymena thermophilia* which is antigenically and morphologically related to the GroEL protein. In view of the cross-reactivity of hsp58 antibodies with a similar-sized protein in mitochondria of various species (*Saccharomyces cerevisiae*, *Xenopus laevis*, *Zea*

mays, and human cells), it is likely that hsp58 is identical to P1. Very recently, Waldinger et al. (22) reported the partial (36 amino acids from the N-terminal end) amino acid sequence of a protein from human lymphocytes which moves in two-dimensional gels in the same position as does the mature P1 protein. The reported sequence shows complete homology with the deduced amino acid sequence of human P1 protein from positions 27 to 62, which indicates strongly that the protein being studied is identical to the P1 protein. The observed homology beginning with residue 27 in P1 supports our contention that the first 26 amino acids in the P1 sequence shown in Fig. 2 and 3 specify a mitochondrial targeting presequence which is cleaved during maturation of the protein. We suggest that P1, because of its *mito*chondrial localization and homology to chaper*onin*s, be named mitonin.

At present, the exact biological role of P1 in mammalian systems is not clear. Although our genetic and biochemical studies (see introductory paragraph and references 4 to 7) point toward its possible involvement in in vivo MT assembly and function, how a protein located in the matrix compartment could interact with MTs remains very puzzling and unclear. Considering its localization in mitochondrial matrix and strong similarity to the chaperonins, a role of this protein in the assembly and function of mitochondrial enzyme complexes (e.g., pyruvate dehydrogenase, ATPase, and α-ketoglutarate dehydrogenase) also seems likely. It should be mentioned that antibodies and T-cell clones reactive with the 65-kDa major mycobacterial antigen are commonly observed in humans infected with either *M. leprae* or *M. tuberculosis* or immunized with the vaccine strain *M. bovis* BCG. The extensive homology of human P1 to this bacterial antigen suggests that immune reaction with P1 could play a role in development of certain autoimmune diseases (10, 21, 23). Therefore, further studies on the cellular function of this highly conserved protein and its possible role in immune and autoimmune reactions are of great interest.

We thank J. D'Agostaro (University of Toronto) and Mike Tyers (McMaster University) for kindly providing cDNA libraries from HL-60 cells, R. A. Young for suggesting possible sequence homology between P1 and the 65-kDa protein of mycobacteria, and D. Stetsko for technical assistance. Data base searches and sequence analyses were performed by using the NBRF Protein Identification Resource (Georgetown, Md.) and the University of Wisconsin GCG software (processed on the National Research Council CND/SND system, Ottawa, Ontario).

This work was supported by research grants from the Medical Research Council of Canada. S.J. and A.K.D. made comparable contributions to this study. The order of appearance of their names was determined by draw.

LITERATURE CITED

1. **Ellis, R. J.** 1987. Proteins a molecular chaperones. Nature (London) **328**:378–379.
2. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods of predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
3. **Goloubinoff, P., A. A. Gatenby, and G. H. Lorimer.** 1989. GroE heat shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. Nature (London) **337**:44–47.
4. **Gupta, R. S., and R. A. Austin.** 1987. Mitochondrial matrix localization of a protein affected in mutants resistant to the microtubule inhibitor podophyllotoxin. Eur. J. Cell Biol. **45**: 170–176.
5. **Gupta, R. S., and A. K. Dudani.** 1987. Mitochondrial binding of a protein affected in mutants resistant to the microtubule inhibitor podophyllotoxin. Eur. J. Cell Biol. **44**:278–285.
6. **Gupta, R. S., and T. K. W. Ho, M. R. K. Moffat, and R. Gupta.** 1982. Podophyllotoxin-resistant mutants of Chinese hamster ovary cells: alteration in a microtubule associated protein. J. Biol. Chem. **257**:1071–1078.
7. **Gupta, R. S., T. J. Venner, and A. Chopra.** 1985. Genetic and biochemical studies with mutants of mammalian cells affected in microtubule related proteins other than tubulin: mitochondrial localization of a microtubule related protein. Can. J. Biochem. Cell Biol. **63**:489–502.
8. **Hemmingsen, S. M., C. Woolford, S. M. van der Vies, K. Tilly, D. T. Dennis, C. P. Georgopoulos, R. W. Hendrix, and R. J. Ellis.** 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature (London) **33**:330–334.
9. **Hoffman, W., J. K. Franz, and W. W. Franke.** 1985. Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis epidermis and evolutionary conservativity of helical and nonhelical domains. J. Mol. Biol. **184**:713–724.
10. **Holoshitz, J., A. Klajman, I. Drucker, J. Lapidot, A. Yaretzky, A. Frankel, W. Van Eden, and I. R. Cohen.** 1986. T lymphocytes of rheumatoid arthritis patients show augmented reactivity to a fraction of mycobacteria cross-reactive with cartilage. Lancet **ii**:305–309.
11. **Huynh, T.-V., R. A. Young, and R. W. Davis.** 1985. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49–78. D. M. Glover (ed.), DNA cloning, vol. 1. IRL Press, Oxford.
12. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8149.
13. **McMullin, T. W., and R. L. Hallberg.** 1987. A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in *Tetrahymena thermophilia*. Mol. Cell. Biol. **7**:4414–4423.
14. **McMullin, T. W., and R. L. Hallberg.** 1988. A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the *Escherichia coli groEL* gene. Mol. Cell. Biol. **8**:371–380.
15. **Mehra, V., D. Sweetser, and R. A. Young.** 1986. Efficient mapping of protein antigenic determinants. Proc. Natl. Acad. Sci. USA **83**:7013–7017.
16. **Schatz, G.** 1987. Signals guiding proteins to their correct locations in mitochondria. Eur. J. Biochem. **165**:1–6.
17. **Schwartz, R. M., and M. O. Dayhoff.** 1978. Origins of prokaryotes, eukaryotes, mitochondria and chloroplasts. A perspective is derived from protein and nucleic acid sequence data. Science **199**:395–403.
18. **Shinnick, T. M.** 1987. The 65-kilodalton antigen of *Mycobacterium tuberculosis*. J. Bacteriol. **169**:1080–1088.
19. **Shinnick, T. M., D. Sweetser, J. Thole, J. van Embden, and R. A. Young.** 1987. The etiologic agents of leprosy and tuberculosis share an immunoreactive protein antigen with the vaccine strain *Mycobacterium bovis* BCG. Infect. Immun. **55**: 1932–1935.
20. **Thole, J. E. R., W. J. Kealen, A. H. J. Kolk, D. G. Groothuis, L. G. Berwald, R. H. Tiesjema, and J. D. A. van Embden.** Characterization, sequence determination and immunogenicity of a 64-kilodalton protein of *Mycobacterium bovis* BCG expressed in *Escherichia coli* K-12. Infect. Immun. **55**:1466–1475.
21. **van Eden, W., J. E. R. Thole, R. van der Zee, A. Noordzij, J. D. A. van Embden, E. J. Hensen, and I. R. Cohen.** 1988. Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis. Nature (London) **331**:171–173.
22. **Waldinger, D., C. Eckerskorn, F. Lottspeich, and H. Cleve.** 1988. Amino acid sequence homology of a polymorphic cellular protein from human lymphocytes and the chaperonins from *Escherichia coli* (groEL) and chloroplasts (rubisco binding protein). Biol. Chem. Hoppe-Seyler **369**:1185–1189.
23. **Young, D., R. Lathigra, R. Hendrix, D. Sweetser, and R. A. Young.** 1988. Stress proteins are immune targets in leprosy and tuberculosis. Proc. Natl. Acad. Sci. USA **85**:4267–4270.